Copyright © 1997 by the Genetics Society of America

Structured Coalescent Processes on Different Time Scales

Magnus Nordborg

Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637-1573

Manuscript received December 10, 1996
Accepted for publication April 24, 1997

ABSTRACT

It is demonstrated that the structured coalescent model can readily be extended to include phenomena such as partial selfing and background selection through the use of an approximation based on separation of time scales. A model that includes these phenomena, as well as geographic subdivision and linkage to a polymorphism maintained either by local adaptation or by balancing selection, is derived, and the expected coalescence time for a pair of genes is calculated. It is found that background selection reduces coalescence times within subpopulations and allelic classes, leading to a high degree of apparent differentiation. Extremely high levels of subpopulation differentiation are also expected for regions of the genome surrounding loci important in local adaptation. These regions will be wider the stronger the local selection, and the higher the selfing rate.

NATURAL selection on the molecular level can be studied indirectly through its effects on variation at linked sites that are not themselves under selection (KREITMAN and AKASHI 1995). While it is possible to model these effects in many different ways, the theory of gene genealogies, or coalescent theory, plays a particularly important role because of its close relationship with actual samples from populations (HUDSON 1990; DONNELLY and TAVARÉ 1995). In this context, the goal is to determine how given selective processes in a population are expected to effect a sample of sequences from that population.

The present article has two purposes. First, I wish to demonstrate that the standard structured coalescent, commonly used to model geographic subdivision and balancing selection, can easily be extended to include ostensibly complex phenomena such as partial selfing and background selection, through the use of an approximation based on separation of time scales. Second, to exemplify the approach, I show how the expected coalescence time for a sample of size two is affected by a combination of background and various forms of balancing selection in a geographically subdivided, partially selfing population. This quantity is of interest because of its direct relationship with commonly used measures of population variability.

The article has the following structure. In the first section, I use the classical two-deme model to exemplify the methods used in the remainder of the article and to remind the reader that models of geographic subdivision can also be used to model the subdivision into *allelic classes* that occurs when selection maintains more than one allele at a locus. Results for these models are provided mainly for ease of comparisons with later results; the only new result in this section is the demonstration that the previously derived expression for the reduction in variability expected under background selection (HUDSON and KAPLAN 1995; NORDBORG *et al.* 1996a) can be obtained from the simple two-deme model.

In the second section, I examine how background selection interacts with balancing selection or geographic subdivision. I also show how balancing selection can interact with geographic subdivision in different ways depending on whether the polymorphism is maintained by local adaptation (*i.e.*, in a cline) or not. The latter model has been studied previously (KAPLAN *et al.* 1991), whereas the former results are new. The third section introduces selfing. It has recently been shown that the coalescent process can be extended to incorporate partial selfing quite easily (NORDBORG and DONNELLY 1997). These results are used here to extend the simple models of the first section to include selfing populations. In the fourth section the results of the preceding three sections are combined to yield a general model.

MIGRATION OR SELECTION

Imagine a standard Wright-Fisher population composed of N diploid individuals, so that there are $2N$ copies of each gene. It is assumed throughout that N is large. The population is subdivided into two classes (*e.g.*, subpopulations) of sizes $N_1 = Nc_1$ and $N_2 = Nc_2$, respectively, where $c_1 + c_2 = 1$. Each of the $2N$ genes can be characterized as belonging to the first or the second class. Let b_{ij}, $i, j = 1, 2$ be the probability that a given gene belonged to class j in the previous generation given that it belongs to class i in the current genera-

Corresponding author: Magnus Nordborg, Department of Ecology and Evolution, University of Chicago, 1101 E. 57th St., Chicago, IL 60637-1573. E-mail: magnus@darwin.uchicago.edu

tion. All parameters are assumed to be constant over time.

The genealogy of a pair of selectively neutral genes sampled from this population (in this context, "gene" simply refers to a nonrecombining piece of DNA) can be described by a discrete-time Markov process with five states: (1, 0), (2, 0), (1, 1), (0, 2), and (0, 1), where (k, l) denotes the state with k distinct genes in the first class, and l distinct genes in the second class. The two genes are "distinct" until their most recent common ancestor has been found. Since a common ancestor will be found eventually, the five states may be partitioned into two equivalence classes: $\mathcal{A} = \{(2, 0), (1, 1), (0, 2)\}$, which is transient; and $\mathcal{B} = \{(1, 0), (0, 1)\}$, which is recurrent.

This process is usually analyzed by assuming that all relevant parameters scale with N as $N \rightarrow \infty$. Formally, it is assumed that the finite limit

$$\lim_{N\to\infty} 2Nb_{ij} = B_{ij}, \quad i, j = 1, 2 \tag{1}$$

exists. Under this assumption, we ignore terms smaller than $O(1/N)$, define $b_{ij} = B_{ij}/(2N)$, and work with the process given by the approximate transition matrix

$$\begin{bmatrix} 1 - \dfrac{B_{12}}{2N} & 0 & 0 & 0 & \dfrac{B_{12}}{2N} \\ \dfrac{1}{2Nc_1} & 1 - \dfrac{B_{12}}{N} - \dfrac{1}{2Nc_1} & \dfrac{B_{12}}{N} & 0 & 0 \\ 0 & \dfrac{B_{21}}{2N} & 1 - \dfrac{B_{12}}{2N} - \dfrac{B_{21}}{2N} & \dfrac{B_{12}}{2N} & 0 \\ 0 & 0 & \dfrac{B_{21}}{N} & 1 - \dfrac{B_{21}}{N} - \dfrac{1}{2Nc_2} & \dfrac{1}{2Nc_2} \\ \dfrac{B_{21}}{2N} & 0 & 0 & 0 & 1 - \dfrac{B_{21}}{2N} \end{bmatrix}, \tag{2}$$

where the states arranged in the order (1, 0), (2, 0), (1, 1), (0, 2), and (0, 1).

The primary significance of assumption (1) is not that it ensures that the b_{ij} are small enough for quadratic terms to be ignored (although this is convenient), but that it ensures that jumps between the two classes occur on the same time scale as coalescent events. Specifically, the expected times between coalescence events and jumps are both of $O(N)$. This fact is used as follows. Assume that the process is in (2, 0), for example. Let $T_{2,0}$ be the random amount of time spent in this state before the process jumps to another state. From the transition matrix (2) we have

$$P(T_{2,0} > t) = \left(1 - \frac{B_{12}}{N} - \frac{1}{2Nc_1}\right)^t. \tag{3}$$

If we measure time in units of $2N$ generations and let $N \rightarrow \infty$, we obtain

$$\begin{aligned} \lim_{N\to\infty} P(T_{2,0} > [2Nt]) &= \lim_{N\to\infty} \left(1 - \frac{B_{12}}{N} - \frac{1}{2Nc_1}\right)^{[2Nt]} \\ &= e^{-(2B_{12}+1/c_1)t}. \end{aligned} \tag{4}$$

In other words, $T_{2,0}$ has an exponential distribution with parameter $2B_{12} + 1/c_1$ in the limit. It is also easy to see from the transition matrix what happens once the process does jump. For example, the process jumps from (2, 0) to (1, 1) with probability

$$\frac{B_{12}/N}{B_{12}/N + 1/(2Nc_1)} = \frac{2B_{12}}{2B_{12} + 1/c_1}; \tag{5}$$

otherwise it jumps to (1, 0). The other states behave analogously, and thus the discrete-time Markov process with transition matrix (2) converges to a continuous-time Markov process where transitions occur as just described. This new, approximate process is generally simpler to analyze than the original, discrete-time process.

In the present article, I will also use an alternative continuous-time approximation to the exact process. Specifically, I will use the limiting process as $N \rightarrow \infty$ without making assumption (1). This *does not* imply that the b_{ij} are large *per se*, only that they are large relative to $O(1/N)$. It *does* imply that jumps between the two classes are much more probable than coalescent events, so that in the $O(N)$ generations that are expected to elapse before a coalescent event occurs, a very large number of these jumps will have taken place. Conversely, in the relatively small number of generations expected to elapse before a jump between classes occurs, a coalescent event is extremely unlikely to have taken place. The consequence of this is that the original process separates into two different processes that occur on different time scales. Transitions from $\mathcal{A}$ to $\mathcal{B}$ (*i.e.*, coalescent events) occur slowly on a time scale that is $O(N)$, whereas transitions *within* $\mathcal{A}$ or $\mathcal{B}$ occur on a much faster time scale. Thus, if we scale time in units of $O(N)$ and let $N \rightarrow \infty$, the individual states in $\mathcal{A}$ and $\mathcal{B}$ will be instantaneous, so that the process will look like an unstructured coalescent process. The rate of this new process, *i.e.*, the rate of jumps from $\mathcal{A}$ to $\mathcal{B}$, is determined by the stationary distribution of the fast process governing transitions within $\mathcal{A}$.

The stationary distribution for two genes can be found directly from the transition matrix governing transitions within $\mathcal{A}$, but the following argument is more illuminating. It is easy to see that a single gene will be in the first class with stationary probability $b_{21}/(b_{12} + b_{21})$ and in the second with stationary probability $b_{12}/(b_{12} + b_{21})$. Since the two genes are independent (to a very good approximation for large N) the stationary probabilities for the states in $\mathcal{A}$ can be found by multiplying the stationary probabilities for each gene. Coalescent events (*i.e.*, jumps to $\mathcal{B}$) can only occur when both genes are in the same class. When both genes are in class i, jumps occur at rate $1/c_i$ per $2N$ generations. Thus the coalescence rate of the new process is

$$\lambda = \frac{b_{21}^2}{(b_{12} + b_{21})^2}\frac{1}{c_1} + \frac{b_{12}^2}{(b_{12} + b_{21})^2}\frac{1}{c_2}$$
$$= 1 + \frac{(b_{12}c_1 - b_{21}c_2)^2}{(b_{12} + b_{21})^2 c_1 c_2} \quad (6)$$

per $2N$ generations.

Alternatively, we can scale time in units of $2N/\lambda$ generations and retrieve the usual coalescent. Since $\lambda \geq 1$, it is clear that coalescent events in the structured coalescent with fast transitions occur at a rate that is greater than or equal to the rate for the unstructured model, with equality if and only if $b_{12}c_1 = b_{21}c_2$, a condition to which we will return shortly. Note that the above argument works without modification if one of the b_{ij} should be $O(1/N)$ or smaller, in which case Equation 6 still holds with the transition probability in question set to zero. In what follows, we will refer to jump processes that obey assumption (1) as *slow* and those that do not as *fast*.

It should be strongly emphasized that the separation of time scales is not restricted to a sample of size 2. As long as the sample size is much smaller than N, any subdivision into classes connected by a fast process results in a process that behaves like the standard, unstructured coalescent on a different time scale.

We now derive the expected coalescence time for two genes. This quantity is directly proportional to several pairwise measures of population variability (HUDSON 1990), *e.g.*, NEI's nucleotide diversity π (NEI 1987) under the infinite-sites model. If jumps between classes are slow, the expected coalescence time for a pair of genes depends on the initial configuration: (2, 0), (1, 1), or (0, 2). Let $T[l]$, $l \in \mathcal{A}$ be the coalescent time for two genes currently in state l. The expectations can be found by conditioning on the first event and utilizing the Markov property to obtain a set of linear equations, *i.e.*,

$$ET[l] = E(\text{time till process first leaves } l) + \Sigma_{m\in\mathcal{A}, m\neq l} P(\text{jump from } l \text{ to } m)ET[m], \quad (7)$$

or

$$ET[(2, 0)] = \frac{1}{2B_{12} + 1/c_1} + \frac{2B_{12}}{2B_{12} + 1/c_1}ET[(1, 1)],$$
$$ET[(1, 1)] = \frac{1}{B_{12} + B_{21}} + \frac{B_{12}}{B_{12} + B_{21}}ET[(2, 0)] + \frac{B_{21}}{B_{12} + B_{21}}ET[(0, 2)],$$
$$ET[(0, 2)] = \frac{1}{2B_{21} + 1/c_2} + \frac{2B_{21}}{2B_{21} + 1/c_2}ET[(1, 1)], \quad (8)$$

which is readily solved to yield

$$ET[(2, 0)] = 1 - \frac{\Delta^2}{\Psi} + \frac{(3\Delta + \Phi)B_{12}^2 c_1}{\Psi(2\Psi + B_{12} + B_{21})},$$
$$ET[(1, 1)] = 1 - \frac{\Delta^2}{\Psi} + \frac{1}{B_{12} + B_{21}} + \frac{\Delta(3\Delta + \Phi)B_{12}B_{21}}{\Psi(B_{12} + B_{21})(2\Psi + B_{12} + B_{21})},$$
$$ET[(0, 2)] = 1 - \frac{\Delta^2}{\Psi} - \frac{(3\Delta + \Phi)B_{21}^2 c_2}{\Psi(2\Psi + B_{12} + B_{21})}, \quad (9)$$

where $\Delta = B_{12}c_1 - B_{21}c_2$, $\Phi = B_{21}c_1 - B_{12}c_2$, and $\Psi = B_{12}^2 c_1 + B_{21}^2 c_2$. The expected coalescence time for a *random* sample of two alleles, ET, is obtained by conditioning on the sample configuration, *i.e.*,

$$ET = \sum_{l\in\mathcal{A}} P(\text{initial state is } l)ET(l) = c_1^2 ET[(2, 0)] + 2c_1c_2ET[(1, 1)] + c_2^2 ET[(0, 2)] = 1 - \frac{\Delta^2}{\Psi}$$
$$+ \frac{2c_1c_2}{B_{12} + B_{21}} + \frac{2\Delta(3\Delta + \Phi)B_{12}B_{21}c_1c_2}{\Psi(B_{12} + B_{21})(2\Psi + B_{12} + B_{21})} + \frac{(3\Delta + \Phi)(B_{12}^2 c_1^3 - B_{21}^2 c_2^3)}{\Psi(2\Psi + B_{12} + B_{21})}. \quad (10)$$

With fast jumps between classes, the initial sample configuration is irrelevant because jumps between the three possible states are instantaneous on the coalescent time scale. ET is obtained immediately as the inverse of the rate parameter λ given by (6), *i.e.*,

$$ET = \frac{1}{\lambda} = 1 - \frac{\delta^2}{\psi}, \quad (11)$$

where $\delta = b_{12}c_1 - b_{21}c_2$ and $\psi = b_{12}^2 c_1 + b_{21}^2 c_2$.

These results are interesting for several reasons. First, we are again reminded of the fact that fast jumps between classes result in a process that behaves like the standard, unstructured coalescent on a faster time scale, so that $ET \leq 1$ unless $\delta = 0$. Second, for the slow process we note that the term Δ^2/Ψ in (10) is independent of the scaling of the B_{ij}. This implies that $\Delta^2/\Psi = \delta^2/\psi$, the corresponding term in (11), and thus, since the remaining terms in (10) are all $O(1/B_{ij})$ or smaller, ET for the slow process converges to ET for the fast process as the B_{ij} become large. At least in this sense, the two approximations seem to overlap smoothly. Third, when the B_{ij} are small, the third term in (10), which is positive, will dominate the expression. It therefore seems as though two different forces are affecting ET: one that increases and one that decreases the expected coalescence time. The former, which can be very strong, is effective only when the rate of exchange between classes is small relative to $1/N$, the latter only when $\delta \neq 0$.

Migration: The above results are perhaps easiest to understand when applied to a simple model of geographic subdivision. Imagine, therefore, a model with two subpopulations of size c_1N and c_2N. The transition probabilities b_{ij}, $i, j = 1, 2$ then correspond to the *backward* migration rates (*i.e.*, the probability that a given gene is an immigrant).

When we interpret the structured coalescent model above as a model of geographic subdivision, the central role played by the quantity δ becomes obvious. We have $\delta = 0$ if and only if

$$b_{12}c_1 = b_{21}c_2, \tag{12}$$

in other words when the number of immigrants is the same in both subpopulations. Since an immigrant in one population is an emigrant from the other, this is equivalent to saying that immigration equals emigration in each subpopulation, so that migration does not affect subpopulation sizes. NAGYLAKI (1980) refers to systems of migration with this property as *conservative* and has shown that, in the strong-migration limit, the effect of nonconservative migration on identity coefficients can be described as a decrease in effective population size. The fast-migration result (6) is a special case of his result. In the remainder of this article, I will extend the use of the term conservative to describe cases where (12) holds even if subdivision is not geographic.

We note that the effect of nonconservative migration in reducing genetic variability is not limited to fast migration. Figure 1 illustrates this by plotting ET against c_1/c_2 and the scaled *forward* migration rate, here assumed to be the same in both directions. It is clear that when the number of migrants is small, ET is sharply increased relative to its neutral value of one, whereas for large numbers of migrants, ET is smaller than one (except when $c_1/c_2 = 1$, which implies conservative migration). The negative effect is stronger the more asymmetric the migration.

Most coalescent models of geographic subdivision have assumed conservative (and often symmetric) migration (TAKAHATA 1988; HUDSON 1990; HEY 1991; HERBOTS 1994). Exceptions include the work of TAJIMA (1989) and of NOTOHARA (1990, 1993a,b), who also looked at the case of fast migration. The results of this section agree with those of the latter two authors.

As shown in the APPENDIX, the expressions for the expected coalescence times under slow migration simplify considerably if migration is assumed to be conservative. They have a particularly simple form when the subpopulations are of equal size (*i.e.*, $c_1 = c_2 = {}^1/_2$) so that the model is completely symmetric. In this case, we have a single migration parameter $m = b_{12} = b_{21}$. It is easy to show that $ET[(2, 0)] \equiv ET[(0, 2)] = 1$,

$$ET[(1, 1)] = 1 + \frac{1}{2M}, \tag{13}$$

and

$$ET = 1 + \frac{1}{4M}, \tag{14}$$

where M is the scaled migration rate $2Nm$ (SLATKIN 1987; STROBECK 1987; TAKAHATA 1988; TAJIMA 1989a; HUDSON 1990; HEY 1991; NOTOHARA 1993a; HERBOTS 1994).

For future reference, we note that WRIGHT's fixation index, F_{ST}, can be calculated approximately from the pairwise coalescence times as

$$F_{ST} = \frac{ET - ET[w]}{ET}, \tag{15}$$

where $ET[w]$ is the average of the expected coalescence times for pairs of genes sampled within a subpopulation (SLATKIN 1991; HERBOTS 1994). For the symmetric model, where $ET[w] = (ET[(2, 0)] + ET[(0, 2)])/2$, we thus have

$$F_{ST} = \frac{1}{1 + 4M}, \tag{16}$$

in agreement with earlier results (WRIGHT 1951; NEI 1975; TAKAHATA 1983; CROW and AOKI 1984; SLATKIN 1991; HERBOTS 1994). Note that this result differs from the classical result $F_{ST} = 1/(1 + 4Nm)$ because the number of demes here is two rather than infinite.

General selection model: Imagine a single population in which a two-allele polymorphism with alleles A_1 and A_2 has been maintained for a long time by some combination of selection and mutation (*e.g.*, heterozygote advantage or mutation-selection balance) at frequencies p and $q = 1 - p$. Let u_{ij} be the mutation rate from A_i to A_j. The forces maintaining the polymorphism are assumed to be strong relative to random drift, so that p and q can be treated as constant to the relevant order of approximation. We are interested in the coalescent process at a neutral locus linked to the locus under

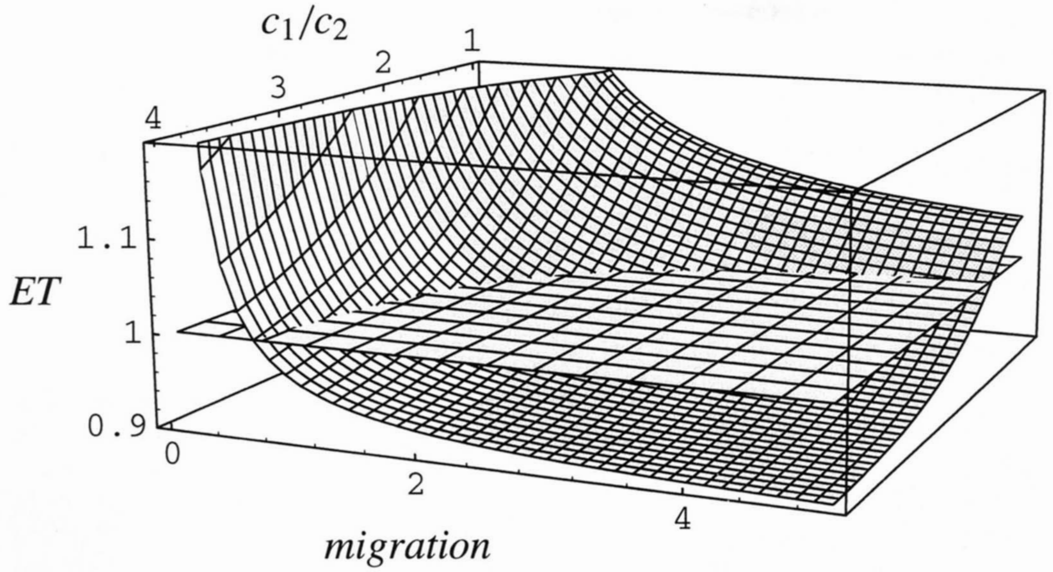

FIGURE 1.—When migration is not conservative, *i.e.*, when (12) does not hold, population subdivision can decrease as well as increase coalescence times. The plot shows *ET*, given by (10), as a function of the ratio of the two subpopulation sizes and the scaled forward migration rate, assumed to be equal in both directions. The plane at $ET = 1$ represents the standard neutral value and is included for comparison.

selection, with recombination rate r. With respect to this locus, the population can be subdivided into two allelic classes, because a given gene is linked either to an A_1 or an A_2 allele. We can thus use the same model as for migration with $c_1 = p$, $c_2 = q$, and transition probabilities given by

$$b_{12} = \frac{q(u_{21} + pr)}{p}, \tag{17}$$

$$b_{21} = \frac{p(u_{12} + qr)}{q}, \tag{18}$$

to linear order in the recombination and mutation rates, as well as in the selection coefficients (HUDSON and KAPLAN 1988; KAPLAN *et al.* 1988; HEY 1991).

Balancing selection: We first look at the case of balancing selection (HUDSON and KAPLAN 1988; KAPLAN *et al.* 1988; HEY 1991). Note that if $u_{ij} \ll r$, we have $b_{12} \approx qr$ and $b_{21} \approx pr$. Since the mutation rate from one functional allele to another is likely to be extremely low, this will be true except for sites that are very tightly linked to the balanced polymorphism.

Assume first, therefore, that mutation is negligible. Under this assumption, $\delta \approx 0$, because recombination, as it is modeled here, is by itself always conservative in the sense of (12). If r is large, so that the fast approximation is appropriate, we thus see immediately from (6) and the accompanying argument that the balanced polymorphism has no effect (to the assumed order of approximation) on the coalescence process. If r is $O(1/N)$, we define

$$\lim_{N\to\infty} 2Nr = R, \tag{19}$$

and obtain the expected coalescence times precisely as for the case of geographic subdivision (the results are given in the APPENDIX). If the balancing selection is symmetric, so that $p = q = {}^1/_2$, we have $ET[(2, 0)] \equiv ET[(0, 2)] = 1$, $ET[(1, 1)] = 1 + 1/R$, and

$$ET = 1 + \frac{1}{2R}, \tag{20}$$

which should be compared with the results for the migration model (HUDSON and KAPLAN 1988; KAPLAN *et al.* 1988; HEY 1991).

If mutation is not negligible compared with recombination, there may be a negative effect on the coalescence time that is not due to linkage, but to nonconservative flow between the allelic classes. Numerical studies indicate that this effect is negligible compared with the effect of linkage except when the b_{ij} are relatively large. For the biological reasons mentioned at the beginning of this section, however, the b_{ij} will never be large unless the u_{ij} are negligible compared with r. This negative effect should therefore never be important under balancing selection.

Background selection: Now assume that the polymorphism is maintained by mutation-selection balance instead of balancing selection. Let A_1 be the wild-type allele and A_2 the (class of) deleterious ones, and define $w_{11} = 1$, $w_{12} = 1 - t_1$, and $w_{22} = 1 - t_2$, where the selection coefficients are assumed to be small as before. The deleterious mutation rate, $u_{12} = u$, is also assumed to be small (although several orders of magnitude greater than u_{12} in the previous section), and the reverse mutation rate, u_{21}, is assumed to be negligible. To the order of approximation, we thus have

$$b_{12} = qr, \tag{21}$$

$$b_{21} = \frac{pu}{q} + pr. \tag{22}$$

Using the well-known mutation-selection equilibrium expression $q \approx u/t_1$ (HALDANE 1927), b_{21} may be approximated by $p(t_1 + r)$, which makes it clearer that its first term is much greater than $O(u)$. Now, to assume that p and q are constant, we must assume that $Nu \gg$

1. This clearly implies that $b_{21} \gg O(1/N)$, so we must use the fast approximation. As we have seen (*cf.* the argument accompanying Equation 6), this is tantamount to saying that the effect of the selected locus on the coalescent at linked sites is equivalent to a reduction in the effective population size. The resulting decrease in expected variability is known as "background selection" (CHARLESWORTH *et al.* 1993). Using (11) we obtain directly, to linear order in q, r, and t_1,

$$ET = 1 - \frac{q}{\left(1 + \frac{r}{t_1}\right)^2}, \tag{23}$$

which is the result found by HUDSON and KAPLAN (1994, 1995) using a related coalescent approach, and by NORDBORG *et al.* (1996a) using diffusion methods.

For n loci in mutation-selection balance that interact multiplicatively, it has been argued that

$$ET \approx \prod_i^n \left(1 - \frac{q_i}{\left(1 + \frac{r_i}{t_{1i}}\right)^2}\right), \tag{24}$$

where q_i, r_i, and t_{1i} are the parameters defined above for each of the n loci; Monte Carlo simulations indicate that this approximation is quite good (HUDSON and KAPLAN 1995; NORDBORG *et al.* 1996a). We note that an induction argument based on the single-locus result in this article also suggests the approximation given by (24), but that the validity of the time-scales approximation as the number of loci increases and the frequency of chromosomes free of deleterious mutations decreases remains to be determined. We will return to this issue in DISCUSSION.

For future use we define α as the ratio of ET under background selection to ET without background selection. Since the latter equals 1 when time is scaled in units of $2N$, α is given by (24). Of course, α is equal to the π/π_0 of CHARLESWORTH *et al.* (1993) and can also be interpreted as N_e/N, where N_e is the effective population size under background selection.

MIGRATION AND SELECTION

To investigate how two different processes, such as migration and selection, jointly affect the coalescent, it is necessary to subdivide the population twice. Imagine, therefore, a Wright-Fisher population of size N diploid individuals, but this time divided into four classes of size N_i, $i = 1, \ldots, 4$. As before, let b_{ij}, $i, j = 1, \ldots, 4$ be the probability that a given gene in class i was in class j in the previous generation. To describe the genealogy of a pair of genes, we now need a minimum of 11 states: the absorbing state plus (2, 0, 0, 0), (1, 1, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1), (0, 2, 0, 0), (0, 1, 1, 0), (0, 1, 0, 1), (0, 0, 2, 0), (0, 0, 1, 1), and (0, 0, 0, 2), where (k, l, m, n) denotes the state with k distinct genes in the first class, l distinct genes in the second class, *etc.*

We are interested only in the special case of this model in which the four different classes are defined by two dichotomous criteria. For example, we may have geographic subdivision into two subpopulations combined with "genetic subdivision" into two allelic classes (Figure 2). This restriction imposes two constraints on the general model. First, the size of each class can be found by multiplying the sizes of its class with respect to the two levels of subdivision (*e.g.*, $N_1 = c_1pN$ in Figure 2). Second, the probabilities of exchange between classes separated by two levels of subdivision (*i.e.*, b_{14}, b_{41}, b_{23}, and b_{32}) will be smaller than the other probabilities.

Three different combinations of time scales are of interest. First, if the probabilities of exchange between the classes are $O(1/N)$ on both levels (*i.e.*, horizontally and vertically in Figure 2), we obtain an approximate process of the same dimensionality, analogous to the slow approximation described above. Second, if the rates of exchange are all high, the process reduces to a single-dimensional one, analogous to the fast approximation. As before, this single-dimensional process will be the usual coalescent on a different time scale. All that is needed to understand this process is thus determining the correct time scale (which may sometimes be algebraically difficult).

The third possibility is that flow across one level of subdivision is fast, whereas flow across the other is slow. This case can be analyzed using an obvious extension of the previously given arguments. Assume (without loss of generality) that jumps between 1 and 2, and 3 and 4 (horizontal flow in Figure 2) occur at a high rate, whereas all other jumps occur with probability $O(1/N)$ or less. Then a very large number of horizontal jumps will occur before any vertical jumps occur, and, on a continuous time scale with time measured in units of $O(N)$, vertical jumps will occur according to the stationary states of three possible "horizontal" equivalence classes, namely

$$\begin{aligned}\mathcal{A} &= \{(2, 0, 0, 0), (1, 1, 0, 0), (0, 2, 0, 0)\}, \\ \mathcal{B} &= \{(1, 0, 1, 0), (1, 0, 0, 1), (0, 1, 1, 0), (0, 1, 0, 1)\}, \\ \mathcal{C} &= \{(0, 0, 2, 0), (0, 0, 1, 1), (0, 0, 0, 2)\},\end{aligned} \tag{25}$$

which of course correspond to the states (2, 0), (1, 1) and, (0, 2), respectively, for the vertical process. Jumps between these states (as well as coalescent events) occur according to an exponential process on a time scale of $O(N)$, as usual. It can be shown that this process has a transition matrix that looks identical to the transition matrix (2), except that B_{ij}, $i, j = 1, 2$ is replaced by $\tilde{B}_{ij}$, where

$$\tilde{B}_{12} = \frac{b_{12}}{b_{12} + b_{21}}(B_{23} + B_{24}) + \frac{b_{21}}{b_{12} + b_{21}}(B_{13} + B_{14}), \tag{26}$$

$$\tilde{B}_{21} = \frac{b_{34}}{b_{34} + b_{43}}(B_{41} + B_{42}) + \frac{b_{43}}{b_{34} + b_{43}}(B_{31} + B_{32}), \tag{27}$$

and c_i, $i = 1, 2$ is replaced by $\tilde{c}_i$, where

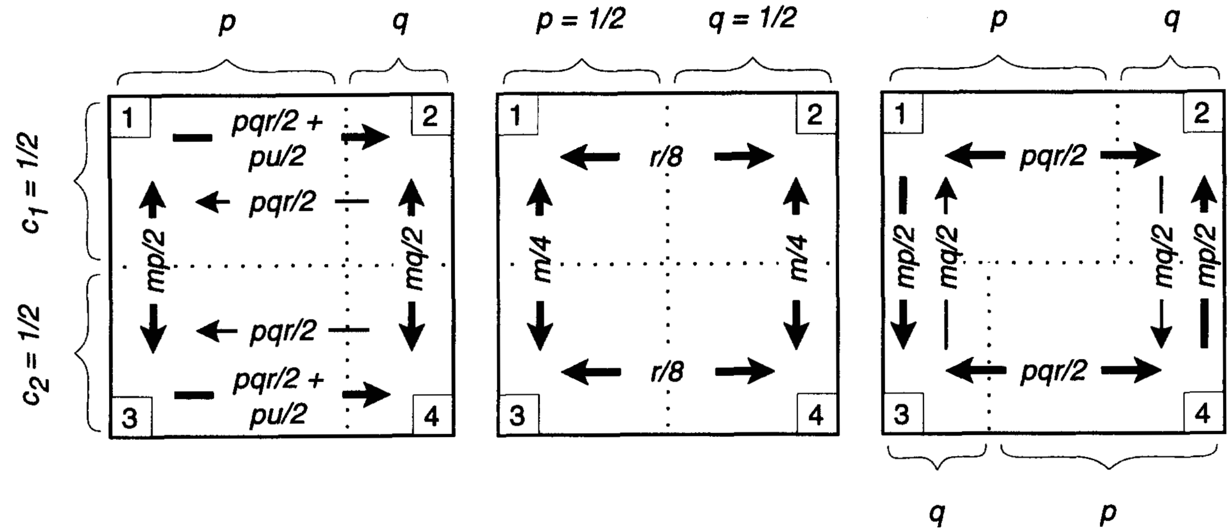

FIGURE 2.—Examples of models with two levels of subdivision. The values on the arrows are $b_{ij}c_i$, *i.e.*, they are proportional to the total number of immigrants; the parameters are defined in the text. From left to right, the models are as follows: migration and background selection, migration and balancing selection without local adaptation, and, finally, migration and balancing selection with local adaptation.

$$\tilde{c}_1 = c_1 + c_2 - \frac{(b_{12}c_1 - b_{21}c_2)^2}{b_{12}^2 c_1 + b_{21}^2 c_2}, \tag{28}$$

$$\tilde{c}_2 = c_3 + c_4 - \frac{(b_{34}c_3 - b_{43}c_4)^2}{b_{34}^2 c_3 + b_{43}^2 c_4}. \tag{29}$$

Since the real subpopulation sizes for the slow, vertical process are $c_1 + c_2$ and $c_3 + c_4$, it is clear that the fast, horizontal process will act to reduce these unless the fast flow is conservative. Coalescent events may therefore occur at a faster rate within each subpopulation. It does not seem possible to say anything equally general about the effect of the fast process on the $\tilde{B}_{ij}$, making the effect on the total coalescence time hard to predict.

General selection-migration model: We begin by combining the models of geographic subdivision and selection introduced above. Let $p_i(q_i)$ be the frequency of $A_1(A_2)$ in subpopulation i. The mutation rates are assumed to be the same in the two subpopulations, but the selection coefficients may differ (which may lead to differences in allele frequencies between the subpopulations). For clarity, we restrict our attention to the case of conservative symmetric migration (*i.e.*, equal subpopulation sizes and a single migration rate m). The critical parameters are the b_{ij}, of which there are now 16 (rather than the two needed for one level of structure). They can be found through standard population genetics theory, but are of course quite complicated in their exact form, and depend on the details of the life cycle assumed. However, using the same approximations as before, it can be shown that, to linear order in migration, selection, recombination, and mutation, the b_{ij} are the elements of the matrix

$$\begin{bmatrix} 1 - \dfrac{p_2 m + q_1(u_{21} + p_1 r)}{p_1} & \dfrac{q_1(u_{21} + p_1 r)}{p_1} & \dfrac{p_2 m}{p_1} & 0 \\ \dfrac{p_1(u_{12} + q_1 r)}{q_1} & 1 - \dfrac{q_2 m + p_1(u_{12} + q_1 r)}{q_1} & 0 & \dfrac{q_2 m}{q_1} \\ \dfrac{p_1 m}{p_2} & 0 & 1 - \dfrac{p_1 m + q_2(u_{21} + p_2 r)}{p_2} & \dfrac{q_2(u_{21} + p_2 r)}{p_2} \\ 0 & \dfrac{q_1 m}{q_2} & \dfrac{p_2(u_{12} + q_2 r)}{q_2} & 1 - \dfrac{q_1 m + p_2(u_{12} + q_2 r)}{q_2} \end{bmatrix}. \tag{30}$$

Migration and background selection: First assume that the polymorphism is maintained by mutation-selection balance with parameters as in the previous section. The deleterious mutations are assumed to act identically in both subpopulations, so the allele frequencies at equilibrium must be equal, as depicted in the left panel of Figure 2. Thus, we have $p_1 = p_2 = p$ and $q_1 = q_2 = q$ in the matrix (30).

As we have seen, background selection must be modeled using the fast approximation. Assume that m is

$O(1/N)$, so that we have the combined fast-slow process we just described. It is obvious that $\tilde{c}_1 = \tilde{c}_2$ and $\tilde{B}_{12} = \tilde{B}_{21}$ because of symmetry. This observation alone gives us $ET[(2, 0)] = ET[(0, 2)] = 2\tilde{c}_i$, and

$$ET[(1, 1)] = 2\tilde{c}_i + \frac{1}{2\tilde{B}_{ij}}, \qquad (31)$$

where the states (k, l) refer to the geographic subdivision. Since the modified subpopulations are of equal size, we also have

$$ET = 2\tilde{c}_i + \frac{1}{4\tilde{B}_{ij}}. \qquad (32)$$

It is easy to show that

$$\tilde{c}_i \approx \frac{1}{2}\left(1 - \frac{q}{\left(1 + \frac{r}{t_1}\right)^2}\right), \quad i = 1, 2, \qquad (33)$$

and that

$$\tilde{B}_{ij} = M, \quad i, j = 1, 2, \qquad (34)$$

where M, as before, is the scaled migration rate $2Nm$. To the order of approximation, background selection affects *only* the effective subpopulation sizes, not migration between subpopulations.

Generalizing to multiple loci in mutation-selection balance, we have $ET[w] \approx \alpha$, and

$$ET \approx \alpha + \frac{1}{4M}, \qquad (35)$$

where α is still given by (24). As a consequence, the *relative* amount of time spent between subpopulations is increased, so that

$$F_{ST} \approx \frac{1}{1 + 4M\alpha}. \qquad (36)$$

Background selection thus leads to an apparent increase in population differentiation, as conjectured by Hudson and Kaplan (1995). For completeness, we note that the case of high migration is trivial (just let $M \rightarrow \infty$ in all expressions).

Balancing selection and background selection: Imagine that the locus under study is linked to a balanced polymorphism on one side with recombination rate r and to a locus in mutation selection balance on the other side with recombination rate r'. It is easy to see that, under the assumptions used throughout this paper (notably the assumption that double recombination events are negligible), this model is identical to the one of the previous section if we replace population subdivision with balancing selection. If the allele frequencies for the balanced polymorphism are equal, for example, we have

$$ET = 1 - \frac{q}{(1 + r'/t_1)^2} + \frac{1}{2R}, \qquad (37)$$

and for the case of multiple loci in mutation selection balance, we would have

$$ET \approx \alpha + \frac{1}{2R}, \qquad (38)$$

as conjectured by Nordborg *et al.* (1996b).

Migration and local adaptation: We now return to the general model of migration and selection and assume that the genetic polymorphism is maintained by some form of local adaptation instead of by mutation-selection balance. To the order of approximation, the resulting model is identical to the more general model of balancing selection in a subdivided population studied by Kaplan *et al.* (1991). These authors did not, however, derive the results for strong local adaptation that will be given below.

Assume that the genetic polymorphism is maintained by local adaptation with negligible mutation between the two alleles. Directional selection favors A_1 over A_2 in the first subpopulation, and A_2 over A_1 in the second. We wish to contrast this situation with one in which the overall allele frequencies are the same, but do not differ between the subpopulations (*i.e.*, there is some form of balancing selection without local adaptation, or the degree of local adaptation is very weak). For simplicity, we will assume that the overall allele frequencies are equal, *i.e.*, $p_1 + p_2 = q_1 + q_2 = {}^1\!/_2$. For this to be the case, the local adaptation must be symmetric so that $p_1 = q_2 = p$ and $q_1 = p_2 = q$, as depicted in the right panel of Figure 2. Note that the case of weak or no local adaptation can be obtained from this model simply by letting $p = q = {}^1\!/_2$, as depicted in the middle panel of Figure 2.

Local adaptation is similar to background selection in many ways. Within each subpopulation, polymorphism is maintained by migration-selection balance, just as it is maintained by mutation-selection balance under background selection. If the migration rate is low, we will have $q \approx m/t_1$, where t_1 is the selection coefficient against heterozygotes (Haldane 1930). Thus, for the allele frequencies p and q to be constant, we need to assume that $m > O(1/N)$, so that only the fast approximation for migration is appropriate [the case of slow migration requires a different argument (M. Nordborg, unpublished data)]. No such restriction applies if the genetic polymorphism is maintained by some other form of balancing selection, nor, obviously, to the rate of recombination.

If recombination and migration are both slow (again, this case is not applicable when the polymorphism is maintained by strong local adaptation), we find the expected coalescence time for each of the 10 possible initial states as before (see Appendix). The mean coalescence time within subpopulations is

$$ET[w, \text{subpopulation}] = 1 + \frac{1}{2R} + \frac{1}{4M + 2R}; \qquad (39)$$

that within allelic classes is

$$ET[w, \text{allelic class}] = 1 + \frac{1}{4M} + \frac{1}{4M + 2R}; \quad (40)$$

and that for a random sample, finally, is

$$ET = 1 + \frac{1}{4M} + \frac{1}{2R} + \frac{1}{4M + 2R}. \quad (41)$$

If recombination and migration are both fast, we use the fast-fast approximation. Because of the high degree of symmetry in the transition matrix, it is possible to find the relevant effective population size explicitly. It can be shown that

$$ET = 1 - \frac{m^2 pq(p-q)^2}{m^2(p^2 - pq + q^2) + 2mpq(p^2 + q^2)r + p^2q^2r^2}. \quad (42)$$

Without local adaptation, $p = q$ and $ET = 1$, whereas if local adaptation is strong we have $m \approx qt_1$, q small, and (42) can be approximated by

$$ET = 1 - \frac{q}{(1 + r/t_1)^2}, \quad (43)$$

which is identical to (23). Thus, local adaptation speeds up the coalescent by decreasing N_e at sites linked to the selected locus, an effect analogous to that of background selection.

This is only true when r is of the same order of magnitude as m, however. For closely linked sites, *i.e.*, when r is $O(1/N)$, we must use the fast-slow approximation instead. Note that, in terms of Figure 2, the vertical process is now fast, and the horizontal one slow, so the indices on the right-hand side of (26)-(29) must be changed in the appropriate manner. This done, the expected coalescence times in terms of the $\tilde{B}_{ij}$ and $\tilde{c}_i$ are the same as for background selection and geographic subdivision. In the present case, however, we have

$$\tilde{c}_i = \frac{1}{2}\left(1 - \frac{pq(p-q)^2}{p^2 - pq + q^2}\right), \quad i = 1, 2, \quad (44)$$

and

$$\tilde{B}_{ij} = \frac{pq}{p^2 + q^2} R, \quad i, j = 1, 2. \quad (45)$$

Thus, if there is no local adaptation, $p = q = {}^1/_2$ and the expected coalescence times are identical to those obtained for balancing selection without geographic subdivision. Under strong local adaptation (*i.e.*, q small), on the other hand, we have

$$ET[(2, 0)] = ET[(0, 2)] = 1 - \frac{pq(p-q)^2}{(p^2 - pq + q^2)} \approx 1 - q, \quad (46)$$

$$ET[(1, 1)] = 1 - \frac{pq(p-q)^2}{(p^2 - pq + q^2)} + \frac{p^2 + q^2}{2pqR} \approx 1 - q + \frac{1}{2qR} \quad (47)$$

and

$$ET = 1 - \frac{pq(p-q)^2}{(p^2 - pq + q^2)} + \frac{p^2 + q^2}{4pqR} \approx 1 - q + \frac{1}{4qR}. \quad (48)$$

Thus, the effective allelic class size is decreased by the migration-selection balance. The major effect of local adaptation, however, is to reduce the effective recombination rate, because heterozygotes are uncommon (KAPLAN *et al.* 1991). We note that the approximations given by (42)-(43) and (48) are consistent with each other, because large R in the latter corresponds to small r in the former, and the approximations converge to $1 - q$.

What about F_{ST} under this model? Because the rate of migration is assumed to be high, one might expect that $F_{ST} = 0$. This is indeed the case when there is no local adaptation ($p = q$) and also when F_{ST} is measured within allelic classes. If F_{ST} is measured in the normal fashion (*i.e.*, without regard to allelic classes), however, we have

$$ET[w, \text{subpopulation}] = p^2ET[(2, 0)] + 2pqET[(1, 1)] + q^2ET[(0, 2)] \approx 1 - q + \frac{1}{R}, \quad (49)$$

so that

$$F_{ST} \approx \frac{1}{1 + 4qR}, \quad (50)$$

where the approximations apply for small q, as before. Thus regions of the genome close to polymorphic sites maintained by migration-selection balance may exhibit extremely high F_{ST} values even though F_{ST} for unlinked sites is zero.

INCORPORATING SELFING

It has recently been shown that the standard coalescent can be extended to incorporate partial selfing by simply keeping track of whether genes are in the same or different individuals and noting that the time spent in states involving two genes in the same individual is negligible on the coalescent time scale (NORDBORG and DONNELLY 1997). This is another example of separation of time scales in the coalescent process. The effect of the mating system is simply to reduce the effective population size to

$$N_e = \frac{2-s}{2} N, \qquad (51)$$

where s is the fraction of offspring produced by self-fertilization, *i.e.*, the selfing rate. This result is in agreement with the classical result (LI 1955; WRIGHT 1969; POLLAK 1987), often written

$$N_e = \frac{1}{1+F} N, \qquad (52)$$

where $F = s/(2-s)$ is the equilibrium inbreeding coefficient for a neutral locus (HALDANE 1924). In the remainder of this section, I will use this insight to show how partial selfing affects models of migration or selection.

Selfing and migration: This case is straightforward. The time until the ancestors of two genes in different subpopulation are found in the same subpopulation is unaffected, but the rate at which two genes in the same subpopulation coalesce is increased by a factor $1+F$, as we just have seen. For the completely symmetric two-deme model, we have $ET[w] = 1/(1+F)$ and

$$ET = \frac{1}{1+F} + \frac{1}{4M}. \qquad (53)$$

Clearly,

$$F_{ST} = \frac{1}{1 + \dfrac{4M}{1+F}}. \qquad (54)$$

Selfing will thus always increase F_{ST}. Note the analogy between the effects of background selection and selfing: both act to decrease N_e, and therefore affect the coalescence times in the same way. Note also that if migration occurs via pollen flow, the migration rate would of course be directly affected by the degree of selfing. This is not the case for diploid migration.

General selection model with selfing: If extending the model of geographic subdivision to incorporate selfing is trivial, extending the model of subdivision into allelic classes is considerably less so. The reason for this is that whereas migration is assumed independent of genotype, recombination is not, because it can only take place in heterozygotes. This is, of course, true with random mating as well, but does not cause a problem because, under random mating, the probability that the parent of a given gene was a heterozygote does not depend on the genotype of the individual in which the gene resides in the current generation. With selfing, this is obviously not true, and we therefore need to divide the population into genotypic as well as allelic classes. For example, a gene in the first allelic class (*i.e.*, linked to an A_1 allele) is either in the A_1A_1 or the A_1A_2 genotypic class. Since, under partial selfing, we also need to keep track of whether two genes are in the same individual or not, 13 states (plus the trivial absorbing state) are needed to model the genealogy of a pair of genes as a discrete-time Markov process. This should be compared with three states under random mating.

Fortunately, it is possible to completely eliminate these extra dimensions by appealing to separation of time scales. As before, coalescent events occur on a time scale that is $O(N)$, and jumps between the allelic classes (*i.e.*, recombination and mutation) occur on a time scale that is either $O(N)$ or much faster. What about jumps between the genotypic classes? These are caused by Mendelian segregation and occur with very high probability per generation. We thus introduce a third time scale to describe these jumps, and simply argue that the process will have reached stationarity with respect to the genotypes long before any other jumps (due to recombination, mutation, or coalescence) take place. Let x, y, and z be the equilibrium frequencies of the three genotypes A_1A_1, A_1A_2, and A_2A_2, respectively. These frequencies will be constant to the assumed order of approximation. The stationary probability that a gene linked to an A_1 allele is in a heterozygote is $y/(2x+y)$, otherwise it is in a homozygote (A_1A_1). The analogous probability for a gene in the second allelic class is of course $y/(2z+y)$. Using these stationary probabilities, and the same approximations as for random mating, the transition probabilities governing jumps between allelic classes can be shown to be

$$b_{12} = \frac{q(u_{21} + p\tilde{r})}{p}, \qquad (55)$$

$$b_{21} = \frac{p(u_{12} + q\tilde{r})}{q}, \qquad (56)$$

where $\tilde{r} = (1-F)r$. These equations should be compared with (17)–(18). Selfing always affects the b_{ij} by reducing the effective recombination rate, and possibly also by altering the equilibrium values of p and q.

Selfing and balancing selection: Given this result, the case of balancing selection is simple. As before, we ignore mutation. Jumps between allelic classes occur as before, but at a rate that is reduced by a factor $1-F$. Coalescence events within allelic classes occur at a rate that is increased by a factor $1+F$. The results obtained for random mating hold with straightforward modifications. For example, in the case $p = q = {}^1\!/_2$, we have $ET[w, \text{allelic class}] = 1/(1+F)$, and

$$ET = \frac{1}{1+F} + \frac{1}{2\tilde{R}}, \qquad (57)$$

where $\tilde{R} = (1-F)R$, as conjectured by NORDBORG *et al.* (1996b).

Selfing and background selection: The case of background selection is also straightforward. A gene is in the class of deleterious alleles with stationary probability

$$\frac{q\tilde{r}}{\tilde{r} + \frac{p}{q}u}, \tag{58}$$

otherwise it is in the class of wild-type alleles (*i.e.*, A_1). Coalescent events occur within these two classes with rates $(1 + F)/q$ and $(1 + F)/p$, respectively. Define $\tilde{t} = (1 - F)t_1 + Ft_2$. Under the assumption that the deleterious allele is rare (*cf.* HALDANE 1927), we then obtain $q \approx u/\tilde{t}$ directly from the genotypic recursions. Using this and the same approximations as before, it is easy to show that

$$ET \approx \frac{1}{1 + F}\left(1 - \frac{q}{\left(1 + \frac{\tilde{r}}{\tilde{t}}\right)^2}\right). \tag{59}$$

This expression should be compared with (23).

Analogously to the case of random mating, we conjecture that the effect of n loci in mutation-selection balance can be approximated by

$$ET \approx \frac{1}{1 + F}\prod_{i=1}^{n}\left(1 - \frac{q_i}{\left(1 + \frac{\tilde{r}_i}{\tilde{t}_i}\right)^2}\right), \tag{60}$$

and redefine α to be N_e/N, where N_e is the effective population size under background selection and selfing (given directly by Equation 60; *cf.* the discussion following Equation 24).

MIGRATION, SELECTION, AND SELFING

This section illustrates how all the forces discussed in this paper act in combination, by extending the model of balancing selection in a subdivided population to include partial selfing and background selection. The results for the simpler models can be recovered as special cases.

No local adaptation: We first look at the case of a balanced polymorphism maintained by symmetric selection acting independently of population subdivision (*cf.* Figure 2, middle panel).

As we have seen, background selection and selfing act to reduce the expected coalescence time within subpopulations and allelic classes. Selfing decreases the rate of exchange between allelic class by a factor $1 - F$, and neither process affects migration. Equations 39–41 thus become

$$ET[w, \text{subpopulation}] \approx \alpha + \frac{1}{2\tilde{R}} + \frac{1}{4M + 2\tilde{R}}, \tag{61}$$

$$ET[w, \text{allelic class}] \approx \alpha + \frac{1}{4M} + \frac{1}{4M + 2\tilde{R}}, \tag{62}$$

and

$$ET = \alpha + \frac{1}{4M} + \frac{1}{2\tilde{R}} + \frac{1}{4M + 2\tilde{R}}. \tag{63}$$

The results for fast migration or recombination can be obtained by letting the relevant parameter go to infinity.

It is easy to show that F_{ST} equals $1/(1 + 4M\alpha)$ for unlinked loci but is affected by R otherwise. The F_{ST} statistic can also be used to with respect to the allelic classes as

$$\frac{ET - ET[w, \text{allelic class}]}{ET}, \tag{64}$$

in which case it, loosely speaking, measures the fraction of the total variability that is due to the division into allelic classes. For high migration, this measure equals $1/(1 + 2\alpha\tilde{R})$, which shows that balancing selection has much greater effect on variability in the presence of selfing and background selection (NORDBORG *et al.* 1996b).

Local adaptation: The local adaptation model (*cf.* Figure 2, right panel) can be treated similarly. For sites that are tightly linked to the selected locus, $ET[w, \text{allelic class}] \approx \alpha(1 - q)$,

$$ET[w, \text{subpopulation}] \approx \alpha(1 - q) + \frac{1}{\tilde{R}}, \tag{65}$$

and

$$ET \approx \alpha(1 - q) + \frac{1}{4q\tilde{R}}. \tag{66}$$

In this case, F_{ST} is approximately equal to

$$\frac{1}{1 + 4q\alpha\tilde{R}}, \tag{67}$$

regardless of whether it is measured with respect to subpopulations or allelic classes (this is so because alleles are so strongly correlated with subpopulation). Since recombination is weighted by a factor $4q\alpha(1 - F)$, which may be very small, it is clear that high F_{ST} values may be expected even for loosely linked sites.

DISCUSSION

This article has demonstrated how arguments based on a separation-of-time-scales approximation can be used to include phenomena such as partial selfing and background selection in the general theoretical framework of the structured coalescent. Not only is it possible to model these phenomena in isolation, it is also possible to combine several models into one for greater biological realism without sacrificing any of the advantages of coalescent modeling (such as suitability for computer simulations). As an example, I calculated the expected coalescence time for two genes in a model that includes background selection, partial selfing, geographic subdivision, and linkage to a balanced polymorphism maintained either by local adaptation or by balancing selection. Below, I discuss the main results and their implications.

Fast processes and N_e: The central point of this article is that any process that effectively subdivides the population into a number of classes connected by fast flows (*i.e.*, transition probabilities much greater than $O(1/N)$ per generation) can be modeled simply as a change in N_e. This result is not limited to a sample size of two. Examples include geographic subdivision with high migration rates, background selection, and partial selfing. Arguments based on separation of time scales in the coalescent have been used before (KAPLAN *et al.* 1991; TAKAHATA 1991), but not presented as a general approach.

Because the concept of N_e is perhaps more commonly used than understood, it is worth belaboring what this result means. By saying that a certain population structure can be modeled as a change in N_e, I mean that an appropriate change in the coalescent time scale will retrieve the equivalent unstructured *coalescent process*, allowing standard analytical results and software to be used (DONNELLY and TAVARÉ 1995; NORDBORG and DONNELLY 1997). This claim is much stronger than, for instance, the statement that the expected *coalescence time* for a sample under a certain model of population structure is equivalent to that of an unstructured population of some effective size (NEI and TAKAHATA 1993).

Nonconservative migration: When the flow in a structured coalescent is nonconservative, expected coalescence times may be decreased by subdivision as well as increased (*cf.* Figure 1). Indeed, if the flow is fast, the expected coalescence time will always be reduced, an effect equivalent to a decrease in N_e (NAGYLAKI 1980; NOTOHARA 1993a).

It is easy to see why population structure can increase the total coalescence time, because two genes in different subpopulations cannot coalesce until a migration event brings them to the same subpopulation, but what is the intuition behind the decreased coalescence time under non-conservative migration? When (12) does not hold, there is a net gene flow from one subpopulation to the other. Looking forward in time, a parent in the "upwind" (net donor) subpopulation is more likely to leave offspring than a parent in the "downwind" (net recipient) subpopulation. Looking backward in time, the ancestor of a gene in the downwind population is more likely to have lived upwind than the other way around. Clearly this may cause a reduction in N_e, and, therefore, shorter expected coalescence times.

In the real world, we would often expect migration to be nonconservative, however, most theoretical analyses of subdivided populations have assumed the simplest version of the finite island model, in which migration is both symmetric and conservative. The results from this model are widely used and quoted, yet their robustness to violations of these basic assumptions do not seem to have been investigated.

Background selection: As we have seen, background selection can be interpreted as a form of nonconservative, fast migration. In this case the heuristic explanation just given works even better because "wind" is explicitly included in the model in the form of the (unidirectional) deleterious mutation rate. It should thus be possible to model background selection simply as a reduction in N_e, but this conclusion is contradicted by the negative values of TAJIMA's D statistic seen in simulations of this process (CHARLESWORTH *et al.* 1993, 1995; HUDSON and KAPLAN 1995): if a reduced N_e was the only effect of background selection, TAJIMA's D should have a mean value of zero (TAJIMA 1989b).

This apparent contradiction is resolved by realizing that the extension of the single-locus result to multiple loci is not rigorous. In particular, a random sample from a population will not be drawn according to the stationary distribution of the fast mutation-recombination process. Under the single-locus background selection model, a randomly sampled gene will be linked to a deleterious mutation with probability q, whereas the stationary probability is $b_{12}/(b_{12} + b_{21}) \approx q/(1 + t_1/r)$. This difference will disappear instantly on the coalescence time scale, and can thus be ignored. With multiple loci, however, it is possible that convergence to the stationary state could be slow enough to be detected, especially for large sample sizes and values of θ, which is precisely the circumstances under which significant negative TAJIMA's D have been observed (CHARLESWORTH *et al.* 1995).

There are further inaccuracies associated with the approximations for multilocus background selection. Monte Carlo simulations have shown that whereas (60) is quite accurate for random mating (CHARLESWORTH *et al.* 1993; HUDSON and KAPLAN 1995), it is considerably less accurate for selfing populations (B. CHARLESWORTH, M. NORDBORG and D. CHARLESWORTH, unpublished data). Furthermore, simulation of models that include balancing selection (NORDBORG *et al.* 1996b) and migration (B. CHARLESWORTH, M. NORDBORG and D. CHARLESWORTH, unpublished data) indicate that the statement that background selection only affects the coalescence time within classes (*cf.* Equations 35 and 38) is only correct to a rough approximation for strong background selection.

Selfing: In contrast to the situation with background selection, extending coalescent models to incorporate partial selfing does not seem to pose any difficulties. Selfing has two distinct effects, one minor, and one major. The minor one is that the effective population size is reduced directly by up to a factor of two because of inbreeding. The major one is that the effective rate of rate of recombination is decreased because of decreased heterozygosity. This latter effect is by far the more important one, because it drastically increases the fraction of the genome indirectly affected through linkage to selected sites. If purifying selection is acting, either in the form of selective sweeps (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989), or in the form of background selection (Charlesworth *et al.* 1993), levels of variability may easily be reduced by several orders of magnitude, dwarfing the twofold reduction due to inbreeding (Charlesworth *et al.* 1993; Nordborg *et al.* 1996b). Similarly, if selection acts to maintain a given polymorphism in a selfing species, a much larger region of the genome will be affected, a point to which we will now turn.

Detecting selection: The action of balancing selection may be detected because it leads to a peak of variability surrounding the selected site (Kreitman and Aguadé 1986; Hudson and Kaplan 1988; Kaplan *et al.* 1988). The results of this article suggest three situations when such a peak may be easier to detect. First, selfing will increase the size of the region affected. Second, any background selection should decrease the variability within each allelic class, making the peak much more apparent. This may be especially relevant in a selfing population where the effect of background selection is expected to be considerable (Nordborg *et al.* 1996b). Third, if a balanced polymorphism is maintained by local adaptation (*i.e.*, in a cline), the peak will also be much wider. Under the right conditions, it may even be possible to scan the genome directly for such polymorphisms.

Apparent subpopulation differentiation: The present results also show that population subdivision may be severely overestimated under some conditions. Because both selfing and background selection decrease coalescence times within subpopulations, they will inflate F_{ST} values. This is, of course, consistent with the notion that F_{ST} depends on $N_e m$, but it nonetheless implies that it is impossible to draw conclusions about *isolation* from this statistic. If, for example, background selection or selective sweeps were to reduce variability by a factor of 100, F_{ST} would behave as if the migration rate was one-hundredth of its actual value.

Another concern is linkage to polymorphisms maintained by local adaptation in a cline. As shown by (67), F_{ST} values will be extremely inflated in a region surrounding such loci, and, especially for organisms with a high natural rate of self-fertilization, this region may be quite wide. Indeed, very high F_{ST} values and extensive allozyme linkage disequilibria have often been found in highly selfing plants (*e.g.*, Hamrick and Godt 1990 and other contributions in the same volume). The results presented here suggest a simple interpretation for these observations.

I thank B. Charlesworth and D. Charlesworth for suggesting the model studied in this article and for continual discussions. I am very grateful to T. Nagylaki for patiently explaining some of the intricacies of migration theory, as well as for extensive comments on the manuscript. The manuscript also benefited from the comments of P. Donnelly, R. R. Hudson and an anonymous reviewer. This work was supported by National Science Foundation (NSF) grant DEB-9217683 to D. Charlesworth and by NSF grant DEB-9350363 to J. Bergelson.

LITERATURE CITED

Charlesworth, B., M. T. Morgan and D. Charlesworth, 1993 The effect of deleterious mutations on neutral molecular variation. Genetics **134:** 1289–1303.

Charlesworth, D., B. Charlesworth and M. T. Morgan, 1995 The pattern of neutral molecular variation under the background selection model. Genetics **141:** 1619–1632.

Crow, J. F., and K. Aoki, 1984 Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. Proc. Natl. Acad. Sci. USA **81:** 6073–6077.

Donnelly, P., and S. Tavaré, 1995 Coalescents and genealogical structure under neutrality. Annu. Rev. Genet. **29:** 401–421.

Haldane, J. B. S., 1924 A mathematical theory of natural and artificial selection. Part II. Proc. Camb. Phil. Soc., Biol. Sci. **1:** 158–163.

Haldane, J. B. S., 1927 A mathematical theory of natural and artificial selection. Part V: Selection and mutation. Proc. Camb. Phil. Soc., Biol. Sci. **23:** 838–844.

Haldane, J. B. S., 1930 A mathematical theory of natural and artificial selection. Part VI: Isolation. Proc. Camb. Phil. Soc., Biol. Sci. **26:** 220–230.

Hamrick, J. L., and M. J. Godt, 1990 Allozyme diversity in plant species, pp. 43–63 in *Plant Population Genetics, Breeding, and Genetic Resources,* edited by A. H. D. Brown, M. T. Clegg, A. L. Kahler and B. S. Weir. Sinauer Associates, Sunderland, MA.

Herbots, H. M., 1994 *Stochastic Models in Population Genetics: Genealogy and Genetic Differentiation in Structured Populations.* Ph.D. thesis, University of London.

Hey, J., 1991 A multi-dimensional coalescent process applied to multi-allelic selection models and migration models. Theor. Popul. Biol. **39:** 30–48.

Hudson, R. R., 1990 Gene genealogies and the coalescent process, pp. 1–43 in *Oxford Surveys in Evolutionary Biology,* edited by D. Futuyma and J. Antonovics. Oxford University Press, Oxford.

Hudson, R. R., and N. L. Kaplan, 1988 The coalescent process in models with selection and recombination. Genetics **120:** 831–840.

Hudson, R. R., and N. L. Kaplan, 1994 Gene trees with background selection, pp. 140–153 in *Non-Neutral Evolution: Theories and Molecular Data,* edited by G. B. Golding. Chapman & Hall, New York.

Hudson, R. R., and N. L. Kaplan, 1995 Deleterious background selection with recombination. Genetics **141:** 1605–1617.

Kaplan, N. L., T. Darden and R. R. Hudson, 1988 The coalescent process in models with selection. Genetics **120:** 819–829.

Kaplan, N. L., R. R. Hudson and M. Iizuka, 1991 The coalescent process in models with selection, recombination and geographic subdivision. Genet. Res. Camb. **57:** 83–91.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking" effect revisited. Genetics **123:** 887–899.

Kreitman, M., and M. Aguadé, 1986 Excess polymorphism at the *Adh* locus in *Drosophila melanogaster.* Genetics **114:** 93–110.

Kreitman, M., and H. Akashi, 1995 Molecular evidence for natural selection. Annu. Rev. Ecol. Syst. **26:** 403–422.

Li, C. C., 1955 *Population Genetics.* University of Chicago Press, Chicago.

Maruyama, T., 1977 *Stochastic Problems in Population Genetics.* Springer-Verlag, Berlin.

Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable gene. Genet. Res. **23:** 23–35.

Nagylaki, T., 1980 The strong-migration limit in geographically structured populations. J. Math. Biol. **9:** 101–114.

Nei, M., 1975 *Molecular Population Genetics and Evolution.* American Elsevier, New York.

Nei, M., 1987 *Molecular Evolutionary Genetics.* Columbia University Press, New York.

Nei, M., and N. Takahata, 1993 Effective population size, genetic diversity, and coalescence time in subdivided populations. J. Mol. Evol. **37:** 240–244.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996a

The effect of recombination on background selection. Genet. Res. Camb. **67:** 159–174.

Nordborg, M., B. Charlesworth and D. Charlesworth, 1996b Increased levels of polymorphism surrounding selectively maintained sites in highly selfing species. Proc. R. Soc. Lond. B **263:** 1033–1039.

Nordborg, M., and P. Donnelly, 1997 The coalescent process with selfing. Genetics **146:** 000–000.

Notohara, M., 1990 The coalescent and the genealogical process in geographically structured populations. J. Math. Biol. **29:** 59–75.

Notohara, M., 1993a The genealogical process of neutral genes with mutation in geographically structured populations. J. Math. Biol. **31:** 123–132.

Notohara, M., 1993b The strong-migration limit for the genealogical process in geographically structured populations. J. Math. Biol. **31:** 115–122.

Pollak, E., 1987 On the theory of partially inbreeding finite populations. I. Partial selfing. Genetics **117:** 353–360.

Slatkin, M., 1987 Heritable variation and heterozygosity under a balance between mutations and stabilizing selection. Genet. Res. Camb. **50:** 53–62.

Slatkin, M., 1991 Inbreeding coefficients and coalescence times. Genet. Res. Camb. **58:** 167–175.

Strobeck, C., 1987 Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. Genetics **117:** 149–153.

Tajima, F., 1989a DNA polymorphism in a subdivided population: the expected number of segregating sites in the two-subpopulation model. Genetics **123:** 229–240.

Tajima, F., 1989b Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Takahata, N., 1983 Gene identity and genetic differentiation of populations in the finite island model. Genetics **104:** 497–512.

Takahata, N., 1988 The coalescent in two partially isolated diffusion populations. Genet. Res. Camb. **52:** 213–222.

Takahata, N., 1991 Genealogy of neutral genes and spreading of selected mutations in a geographically structured population. Genetics **129:** 585–595.

Wright, S., 1951 The genetical structure of populations. Ann. Eugenics **15:** 323–354.

Wright, S., 1969 *Evolution and the Genetics of Populations*, Vol. 2. University of Chicago Press, Chicago.

Communicating editor: R. R. Hudson

APPENDIX

Coalescence times with conservative migration: If (12) holds, the expected coalescence times (9) become

$$ET[(2, 0)] = 1 + \frac{(c_1 - c_2)c_2}{1 + B_{12}c_1 + B_{21}c_2},$$

$$ET[(1, 1)] = 1 + \frac{1}{B_{12} + B_{21}},$$

$$ET[(0, 2)] = 1 + \frac{(c_2 - c_1)c_1}{1 + B_{12}c_1 + B_{21}c_2}, \qquad \text{(A1)}$$

and (10) can be written

$$ET = 1 + \frac{2c_1c_2}{B_{12} + B_{21}} + \frac{c_1c_2(c_1 - c_2)^2}{1 + B_{12}c_1 + B_{21}c_2} \qquad \text{(A2)}$$

(Takahata 1988; Tajima 1989a; Hey 1991; Notohara 1993a; Herbots 1994). Note that the expected coalescence times for samples from within a subpopulation (*i.e.*, $ET[(2, 0)]$ and $ET[(0, 2)]$) are affected by subdivision except when the subpopulation sizes are equal (Herbots 1994), but that $c_1ET[(2, 0)] + c_2ET[(0, 2)] = 1$, in agreement with classical results (Maruyama 1977).

Coalescence times with balancing selection: The expected coalescence times (9) become

$$ET[(2, 0)] = 1 + \frac{(p - q)q}{1 + 2pqR},$$

$$ET[(1, 1)] = 1 + \frac{1}{R},$$

$$ET[(0, 2)] = 1 + \frac{(q - p)p}{1 + 2pqR}, \qquad \text{(A3)}$$

and (10),

$$ET = 1 + \frac{2pq}{R} + \frac{(p - q)^2pq}{1 + 2pqR} \qquad \text{(A4)}$$

(Hudson and Kaplan 1988; Kaplan *et al.* 1988; Hey 1991).

Coalescence times with balancing selection and migration: Because of the high degree of symmetry in the absence of local adaptation, it is easy to find the expectations by the usual method of conditioning on the first event. We obtain $ET[(2, 0, 0, 0)] = ET[(0, 2, 0, 0)] = ET[(0, 0, 2, 0)] = ET[(0, 0, 0, 2)] = 1$, and

$$ET[(1, 1, 0, 0)] = ET[(0, 0, 1, 1)]$$
$$= 1 + \frac{1}{R} + \frac{1}{2M + R},$$

$$ET[(1, 0, 1, 0)] = ET[(0, 1, 0, 1)]$$
$$= 1 + \frac{1}{2M} + \frac{1}{2M + R},$$

$$ET[(1, 0, 0, 1)] = ET[(0, 1, 1, 0)]$$
$$= 1 + \frac{1}{2M} + \frac{1}{R}. \qquad \text{(A5)}$$